Chapter 1

Searching the Web

Web searching has become a part of the day-to-day life of any person who has a computer. Searches vary from words, phrases, products, people, images, and many other forms. Today, when you want to know something you simply do a Web search. There are many search engines but none, according to `searchengineshowdown.com`, quite as large due to Google's inclusion of file types such as PDF, DOC and PS, not indexed by other search engines. What makes Google so powerful is not just the sheer number of files it has indexed, but the tools that make arriving at the answer to your question simple and straightforward.

This chapter helps you learn to better form your queries and introduce you to the basic Google search. Even the basic Google search is powerful and gives you an amazing amount of control over how you view your results and how to best focus your search so that you arrive quickly and easily at the results you seek.

IN THIS CHAPTER

- Searching the Web
- How Google searches the Web
- Using Keywords
- The Search Result Page
- Using the I'm Feeling Lucky Button
- Getting the Right Pages
- Setting Search Preferences

Forming the Basic Query

The most basic Google search is performed by navigating to the Google home page at `www.google.com`, typing a search term into the Google Search box (shown in Figure 1.1), and pressing Enter. Before you can say Google, your results appear for you to peruse.

Your search results appear with a short description of the resulting page and a link to navigate to it. Clicking the link causes you to navigate away from Google to the page. To return to Google, click the Back button in your browser. This is often represented as a left-facing arrow icon.

COPYRIGHTED MATERIAL

FIGURE 1.1

Type a search term and press Enter to find out almost anything.

Keywords

The search terms you type on the Google home page are called *keywords*. When typing keywords, there are a few things to remember. More keywords narrow your search, but too many may cause you to be too specific and miss the information you are looking for. Google pays attention to whether the keywords are singular or plural and also searches based on the order of your keywords. For example, if you search on the two keywords town and dog in that order you will get results that start off with links to dogtown.com, a skateboard company. But the second link is to a pet shop. Change the order to dog and then town and the results change. Now at least the first 15 or so results refer specifically to Dogtown skateboarding.

Google also uses something known as *stemming*. This means that Google automatically looks for variations of a keyword. For example, if type the search text **cloning your cat**, Google also looks for clone cat (see `www.savingsandclone.com`).

TIP **Don't use the word *and* as a keyword. Google will ignore it. Google automatically looks for all of your keywords within a page. The most relevant page will contain all or most of the keywords.**

Occasionally you will want to create phrases out of your keywords. This is particularly useful when searching on names. To create a phrase, enclose a group of words in quotes like this: "dog town". If you try this search you will see that it gives you different results from the previous two searches on those terms. If you search on the terms John and Smith, Google looks for pages that contain the name John and the name Smith, not necessarily in the same name. The resulting Web page may have the names John Anderson and Joan Smith. Putting the name John Smith in quotes like this — "John Smith" — forces Google to search for that phrase specifically.

Querying the results

One thing Google is good at is giving you as many possible results for your query as possible. This is good in that Google is thorough and bad in that too many results could mean that the result you are really looking for is lost in the weeds. Google gives you the opportunity to do a second search but only through the results of the previous search.

Let's say you have searched on dog breeds. The results are innumerable (actually just over 11 million). This is far too many pages to search manually. At this point you have two options: Go back and form a more specific query, or search through the 11 million results to narrow the number. At the bottom of the results page you will see a link, Search within results. Clicking this link will cause a new Search within results search box to appear. Enter a new search term. Continuing this example, try typing **"fox terrier"**. Even though this example returns 780,000 results, they are more specific to the dog breed fox terrier. Unfortunately you cannot continue searching through your results a second time. If there are still too many results and not specific enough you will have to go back and form a better query. Knowing how Google does is job can help you form better queries.

How Google Searches the Web

To get the best out of Google it helps to know a little bit about how Google goes about searching the Web. This will help you create better queries and understand the results you are seeing and why they are ordered the way they are. Google ranks the results of a search by using "traditional" factors such as the URL (Web page address such as `www.google.com`), *meta tags* (invisible pieces of information about a Web page added by a Web page author), keywords (search terms), and its own patented technology called PageRank.

Google follows four steps to complete the search:

- Finds all the pages that match the keywords on the page.
- Ranks the pages using "traditional" factors (URL, meta tags, and keyword frequency).
- Calculates the relevancy of the link text. How related are the keywords to what appears in the link?
- Google displays the results using PageRank to determine the result order.

Google employs search bots (Web crawlers), which are special programs that search through Web pages, evaluating the pages based on certain criteria and creating an index that allows for rapid search results.

Traditional factors

There is certain information within a Web page that Web crawlers look at to analyze where a page should fall within the results of a search. The term for this is *relevancy*. Pages in a search engine most often appear in the results based on the most relevant pages first. There are several factors in determining the degree of relevancy a page might have. Some of these are obvious, such as the keywords appearing within the text of the Web page. Other, less obvious, factors are involved in determining the relevancy of the page. These include meta tags, descriptive tags, and link text.

Meta tags

Meta tags refer to information that is placed within the header of a Web page that is not visible to the person viewing the page. The information in these tags, which is stored in name/value pairs, is passed to search bots or Web crawlers like the ones Google uses. The information stored in these tags often includes keywords and descriptions of the page that the Web page author would like to associate with the page.

Even though Google does not give these tags much weight, it still looks for keywords in them. There are many types of meta tags but the two most commonly used for assisting in PageRank are the `<DESCRIPTION>` tag and the `<KEYWORD>` tag. The `<DESCRIPTION>` tag includes a text string describing the contents of the page. This is the description used by Google and other search engines to describe the page to Web searchers. The `<KEYWORD>` tag contains a comma-delimited list of keywords the Web page author feels are important in finding the Web page. This includes terms that may not be found within the actual text of the page. For example, an e-commerce site might include keywords such as *hefty man* in the `<KEYWORD>` tag but not include this

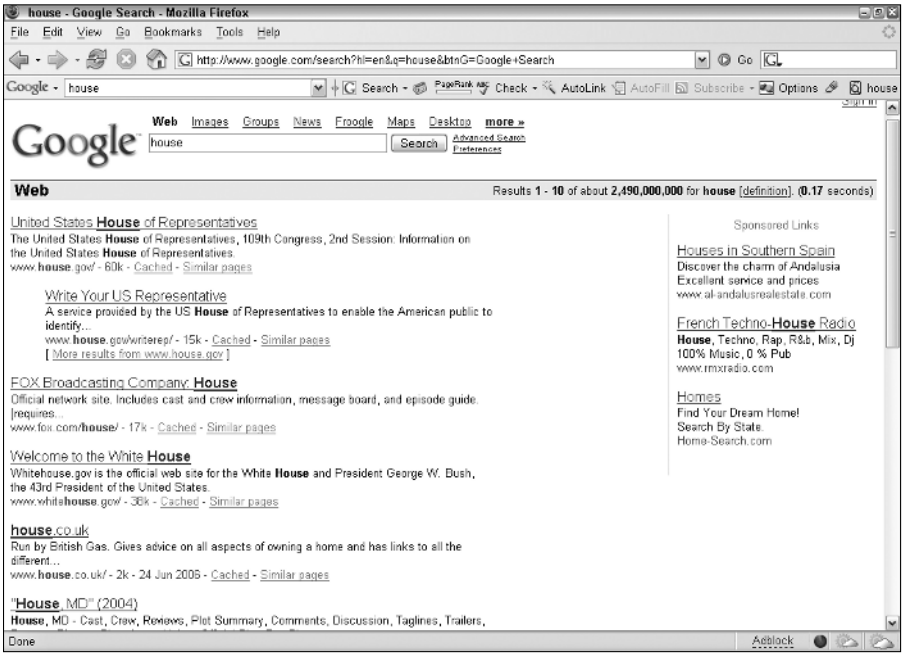
term in the actual text viewed on the Web site. Someone then searching for hefty man clothing may then find this page. Where, without this keyword the page may not have been included in their search results. Other HTML tags, such as <TITLE>, <H1>, and <H2> are also searched by Google's Web crawler for keywords.

The link text

The URL, or link text, has always been important in determining a page's relevance. When the URL of a Web page contains the keyword it will be considered much more relevant by Google than a page that may contain the keywords within the text but not within the link text. For example, when you search for the word *house* using Google, the first five results have the word *house* in their URLs as shown in Figure 1.2.

FIGURE 1.2

When search terms appear in the URL, they are considered more relevant.



CROSS-REF **The Web browser shown in Figure 1.2 displays the Google Toolbar. See Chapter 31 to learn more about downloading and installing the Google Toolbar. Another factor Google takes into account when considering page relevancy is the frequency with which the keyword appears on the Web page. The more times the word you are searching for appears within the Web page the higher the relevancy of that page.**

PageRank

PageRank was created at Stanford University in 1985 as part of a research project studying a new kind of search engine. The project was developed by Larry Page and Sergey Brin. They created a functional search prototype that they called Google. Shortly after creating the PageRank technology Larry and Sergey founded Google, Inc., making use of the PageRank technology as a key element in its new Web search software.

If you search for PageRank on Google, you find the following description:

PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important."

The PageRank that a Web page receives is determined by the number of links that are pointing to the page. So each link is basically a vote but since Google is not a democratic country, not all votes are equal. Some votes are more important and have greater value. PageRank gives a little more value to the votes on pages that are themselves listed higher in the PageRank.

NOTE **Some votes can count against you. Webmasters, in order to increase their PageRank, add their pages to link farms. These are locations on the Internet that will add links to your page. Google punishes this behavior by removing the page from the Google index.**

Of most importance to PageRank is the link to a Web page. The type of Web sites that link to a page will most determine the PageRank. For example, when there is a Web page that talks about Labradors (breed of dogs) and the American Kennel Club (`www.akc.org`) has a link to that page, this link is given greater weight than perhaps a link from a personal Web page. This is because the AKC Web page is specifically about dogs and breeds and is likely to have a very high PageRank of its own.

More About Keywords and Queries

Now that you know a little bit more about how Google searches through Web pages, determines relevancy, and PageRank, it's time to understand a little bit more about the most fundamental part of doing a Web search, the *keyword*.

A keyword is a significant or descriptive word that defines the topic on which you are searching. In other words, keywords define the core of your search. To best find content on the Web you must use the most efficient keywords. Using more keywords narrows your search but doesn't necessarily make it more efficient. You are limited to using a maximum of ten keywords. When you narrow your search, fewer search results are returned, but you may make the search so narrow that you miss what you were looking for. It is sometimes better to group keywords into a search phrase, as described in the first part of this chapter or by creating a search expression.

Introducing search expressions

A *search expression* is a set of keywords and operators such as *and, or,* +, and –. Using search expressions can often help you narrow your search most effectively. Here is a sneak peak at the more detailed discussion in Chapter 2. Use the + sign before any word to tell Google that this word must be in the results. This can be particularly important because Google often leaves out common words that it considers unimportant to the meaning or context. Common words such as who, what, when, where, and how are often ignored when included as keywords. When a common word must be considered, precede it with the plus sign like this: **+how +now brown cow.**

Without the plus signs, Google may have ignored the keywords *how* and *now.*

TIP **You can forget about capitalizing words. Google ignores all capitalization.**

Another useful way to affect the way Google understands keywords is to tell Google what to ignore. Precede keywords that you specifically don't want to see in your results with a – (minus sign). For example, if you are searching on old *M.A.S.H.* TV shows and using the keyword *potter,* and don't want to see an endless stream of pages about Harry Potter, you can precede the word *harry* with a minus sign like this: **–harry potter mash tv**.

Whether using the + or – sign to affect how Google handles your keywords, make certain there is a space before either of these characters when following another keyword.

The query

Type your keywords into the Google Search box as shown in Figure 1.3. You can begin your search by pressing Enter or by clicking Search.

TIP **Don't be afraid to get extra help by clicking the Search Tips link at the bottom of each Google results page.**

Google Search boxes have become ubiquitous. They are everywhere. Google has given the ability to include its search box on any Web page that wants to have one. You don't necessarily have to be on the Google home page to perform your query. Be advised that many of the Google Search boxes that appear on other sites have the ability of searching the entire Web or just that site. Normally, the Web page informs you of any search limitations.

WARNING **Google searches are not secure. Never enter personal or financial information into a Google Search box, particularly one that is on a page other than the Google home page.**

FIGURE 1.3

Type your search term into the Google Search box.

The missing word query

Earlier in this chapter you learned how to create a phrase by enclosing keywords within quotes. Another useful tool is the * (missing word wildcard), which is used only within phrases. This is particular great when you are searching for things like song lyrics and can't remember all the words. For instance, "If you could * my mind love" as a search phrase will return the words from the Gordon Lightfoot song that begins, "If you could read my mind love."

Using the missing word wildcard is also great when you need answers. Here is an example: "the circumference of the Earth is * miles" returns "the circumference of the Earth is 24901.55 miles in the first result." Of course, you could have typed the phrase "the circumference of the Earth is" without the wildcard and the word miles and you might receive different results including the circumference in kilometers or feet. It is a little-used feature but can help you narrow your search.

One last useful feature of the missing word wildcard is that Google does not count this wildcard when calculating whether or not you have exceeded the ten-keyword limit. You might find it useful when too many search terms have become a problem.

The Search Results Page

Once you click the Google Search button, Google launches a new page displaying the results of your search. The first thing you notice in the results page is the number of results Google found in response to your query. For example, if you searched on *dogs,* the page tells you that it is displaying the first 10 results of about 183 million links for dogs. Next to the number of results there is a link to read the definition of dog. To the right of the definition link Google tells you how long it took to perform the search. Perhaps this is bragging rights but I can see no real purpose for this metric.

Search results

Each search result displays the name of the Web page, a short description of the page contents, the link you will navigate if you choose that result (the Web page address), the size of the page, the cached version of the page, and another link called Similar Pages.

Occasionally you may click a link attempting to see one of the results and find that it no longer exists, or that the information on the page has changed since Google last indexed it and the information you are interested in no longer appears on the page. Google keeps a snapshot of the indexed page in a cache (historical storage space) on the Google servers. The cached version of the page is what Google uses to set the PageRank for your query. Use the Cached link when you are unable to access the page containing the information in which you are interested. When the cached link is missing it's because Google has either not yet indexed the page or the owner of the page requested that the page's content not be cached. This often happens with commercial content for which the owner of the Web page expects you to pay.

When you see results that are indented, this means that Google has found your keywords within other pages from the same Web site. Remember your results will not necessarily appear on the site's home page. Clicking the links in the indented results help you jump to different pages within this Web site. You may not find exactly what you are looking for on the page that contains your keywords. You may want to navigate to other pages in the resulting Web site. Navigating through the site can often be challenging. Some Web sites don't provide a menu or simple mechanism for navigating the site. In those cases it is often easiest to start at the site's home page. After navigating to the resulting Web page, look at the URL in the address bar of your Web browser. Copy only the domain name portion, for example, `www.google.com`, and forget everything that appears after the .com, .org, or whatever suffix might appear in the domain name. Paste this back into the address bar of your Web browser and press Enter. That should take you to the site's home page.

Oops

When you receive the result that states: Your search – keyword list – did not match any results, it means that your keywords or phrases matched no Web pages within the Google index. That doesn't mean they don't exist, they have just not been indexed by Google's Web crawlers. Here are three good suggestions:

- Make sure all words are spelled correctly.
- Try different keywords.
- Try more general keywords.

Google is really good about catching misspelled words. Unless the keyword is very obscure, Google usually suggests an alternate spelling. Trying different keywords or more general keywords is probably your answer. If you are have used quotes to create an exact phrase, try doing the search without the quotes.

On to the next one

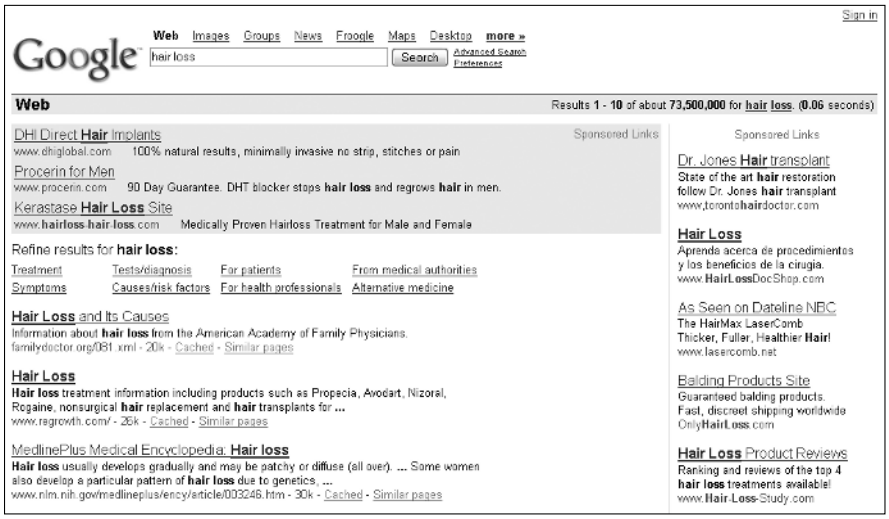
By default Google displays ten results per page. You learn how to adjust that later in this chapter. Once you review all the results in the page you can continue on to the next page by clicking several different things at the bottom of the results page. You can click the blue right arrow at the end of the word *Goooooooooogle*. You can click the Next link beneath that. Last, you can click a specific page number to go directly to a specific page. Ten pages are there initially to choose from. When there are more than ten pages of results, which is often the case, navigating to some page forward causes Google to show you the next ten pages after that. For example, navigating to page 10 causes Google to list pages 1 through 19. Once you navigate away from the first page, a Previous link appears to allow you to navigate backward through the result pages.

Sponsored links

When viewing the results you will notice that on the right side of the page there are *sponsored links*. These links are paid for by companies so you can see them when you search on a specific topic. Some sponsored links may also appear above the other results, and appear with a shaded background to call attention to them. Google also labels them as sponsored links (see Figure 1.4).

FIGURE 1.4

Sponsored links appear as the first results and to the right side of the results page.



Results that assist you

Depending on the keywords you type, Google may present the results in various ways. For example, if you search for golden retriever the results page begins by showing you three images of golden retriever puppies. A similar thing happens when you search on orchids: You are shown three images of orchids.

If you search for a place like Los Angeles or Tampa, the first result is a link to a map of that location. You are then asked if you want to refine your search and look for: dining guides, attractions, lodging guides, shopping, suggested itineraries, and tour and day trips.

Google includes a spell-check feature. When Google thinks you might have misspelled a word, an alternate suggestion is displayed as the first result. If you type **goldon retreiver** as the keywords in the Google Search box the results show a legend "Did you mean:" and then a link to golden retriever, which is the corrected spelling. If you decide that Google is right in its assumption that the keyword(s) is misspelled, you can click the link and the search is performed again with the correct spelling. A new results page appears.

It makes no difference what your search topic is, Google displays within the first links of the results page the links that will help you continue your search.

Using the I'm Feeling Lucky Button

The I'm Feeling Lucky feature is one of Google's unique attributes. What this feature does is bypass the search results page and takes you directly to the first Web page that Google returns from your query.

For example: When you type **golden retriever** and click I'm Feeling Lucky, Google points your browser to the American Kennel Club's Golden Retriever page, where you find information about this breed.

TIP **Use the I'm Feeling Lucky button only for very specific topic searches.**

You can find the I'm Feeling Lucky button on Google's home page. You can also find it when using the Google Toolbar by clicking the little down arrow next to the keyword search box to open the drop-down menu. You can then select I'm Feeling Lucky, as shown in Figure 1.5. The browser navigates to the Web page that most closely matches your keyword search.

FIGURE 1.5

Select I'm Feeling Lucky from the drop-down menu on Google's Toolbar.

Getting the Right Pages

Getting to the right page will be influenced by which keywords you select or your search expression. Because Google returns results that match any of the keywords in your query, a good search expression with the right keywords will help you focus your search.

There are four questions you should ask yourself before doing your search. If you keep the answers to these questions in mind when you search, the results will be much closer to those you were looking for:

- What are you looking for? Conceptualize your search by finding the right keywords or create the right search expression.
- Do your keywords have synonyms or related terms? If so, you might want to think about using them.
- Does your search expression have the correct syntax and use the right search operators?
- After your search, were the results relevant to your query? If not, go back and try refining your keywords and search expression in step 3.

Additional features that can assist you in focusing your search are the ability to search similar pages and restricting your search to a single site.

- **Similar Pages:** Google displays a link under each search result called Similar Pages. This link takes you to pages that are relevant to your search result and most like that particular result.
- **Site Search:** You can restrict your search to a single site or domain. Type your keywords, then type the word **site:**, making certain to include a colon, followed by the site address — for example, **firefox site:mozilla.com**.

Setting Search Preferences

There are five global preferences you can change. These include

- Interface Language
- Search Language
- Safe Search Filtering
- Number of Results
- Results Window

If you go to the Google main page (`www.google.com`) you see three links next to the box where you type the search expression. One of these is the Preferences link. Clicking this link takes you to a page where you can modify the way Google delivers its search results.

Changing your preferences will apply to all Google services.

Setting languages

The first preference, Interface Language, allows you to select from among the 119 different languages of which Google is capable to display tips and messages. All Google pages will be displayed in the language you select, not including Google search results.

See Chapter 23 to learn more about setting language preferences in Google.

The Search Language preference allows you to choose which language Google uses to perform its Web page search. Results will include Web pages written in your selected language. There are two options: You can search for pages written in any language, or you can search only for pages written in the languages you

select. You are provided with a list of 31 languages, and you can select as many as you want. Click the checkbox next to the language to select it. Google recommends choosing the first option: Search for pages written in any language. This is recommended because when you do your search you will get more results. If you only choose a few or a single language only, Google will not search the whole World Wide Web; it searches only the pages written in the specified language or languages.

Setting filter options

SafeSearch Filtering filters adult sites containing explicit sexual content. There are three SafeSearch options:

- **Use strict filtering:** This eliminates search results with explicit images and explicit textual content.
- **Use moderate filtering:** This filters only explicit images. This is the default setting for Google.
- **Do not filter my search result:** No filtering takes place.

SafeSearch filtering its not 100 percent accurate but it does a very good job. Chapter 4 contains much more information about SafeSearch.

Result number setting

When searches return a large number of results it can be time consuming to continue paging through the results. Instead of using the default number of ten results per page you can choose to customize the results page by choosing a larger number of results per page. To change the number of results that Google displays use the Number of results preference setting. You may choose to see 10, 20, 30, 50, or even 100 results for page by selecting from the drop-down list.

NOTE **The more results you choose, the longer it takes the results to load into your browser.**

Opening a new results window

The last preference is Results Window. By selecting this option you can choose to view search results in a new browser window. Every time you do a search and click on a search result link a new browser window opens on your computer displaying the results of your search. This is particularly handy if you want to compare or view many results at the same time.

Another option to seeing your results displayed in a new window is after you perform your Google search, you have the option of opening each result in a new window by holding the Ctrl key and clicking the link with your mouse. On a Macintosh this is the Option key.

Prefetching

Prefetching is a special feature of Mozilla and Firefox browsers. Similar to I'm Feeling Lucky, Prefetching bets that what you are looking for is in the first Web page. With Prefetching set in your browser, Google begins downloading the first Web page that appears on the Google search results page. This increases efficiency as the first page is already loaded and ready to display rather than waiting for it to load after clicking the link in the search results.

CAUTION **If you have this feature turned on in your browsers, you may end up with cookies from Web pages you have never visited along with additional pages in your Firefox or Mozilla caches. If these pages contain images, they will be saved in your Temporary Internet files.**

Summary

Google is much more than simply a search engine. It is an information resource. Still, what makes Google what it is, is the ability to search the Web and get the answers you are looking for. In this chapter you learned about searching with keywords and phrases, received an introduction to search expressions, and learned just how Google comes up with its results.

Searching Google and getting to the answers is a bit of an art, but luckily all arts can be learned through practice. This first look at Google searching may have seemed simple. That's because searching with Google is simple. There are ways to create complex queries. The next chapter goes into greater detail in describing how to create a search expression and using the advanced search features of Google.